8-6-2021

Binge-watching and the Spacing Effect

Michael R. Austin
James Madison University

Binge Watching and the Spacing Effect

Michael R. Austin

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

August 2021

FACULTY COMMITTEE:

Committee Chair: Kethera Fogler

Committee Members/ Readers:

Jeff Dyche

Jessica Irons

Table of Contents

List of Tables

List of Figures

Abstract

Binge-watching, defined as consuming at least three episodes or three hours of video media in one sitting, is an increasingly prevalent behavior in the digital age. But scant research exists investigating how binge-watching affects memory for what was watched. Literature surrounding the spacing effect, defined as superior memory for information presented repeatedly across longer spans of time, would predict a memory deficit for binged material. However, findings from previous unpublished research by Fogler and colleagues do not align with this prediction. To investigate the dissonance, the aim of this research is to replicate and extend the work of Fogler and colleagues, addressing the question: Does binge-watching result in poorer memory for what was watched? University students ($n = 92$) were randomly assigned to a group that either 1) watched three episodes of the Netflix series, House of Cards, in one three-hour long session (massed) or 2) watched these same episodes spaced out over a three-week period (i.e., one episode a week; spaced). After a retention delay of either one week or four, participants' memory of episode content was tested using the Binge-Watching Episode Assessment. A 2x2 factorial ANOVA was conducted to investigate the effect of viewing condition (spaced, massed) and delay (one-week, four-week) on memory. Results indicate no consistent benefit of spacing out episodes. At a delay of four-weeks, binge-watchers actually had substantially better memory than their spaced counterparts. Consistent with deficient processing accounts, elevated attention for binged content is discussed as a possible mediator.

Introduction

Binge-watching (or 'binge-viewing' or 'television marathoning') can be loosely defined as consuming a large quantity of video media in one sitting. The quantity can be accounted for either in number of episodes watched, or time spent watching. Three episodes, or three hours watched is a typical marker used to indicate binge-watching behavior (Pittman & Steiner, 2019; Rubenking & Bracken, 2018; Fogler et al., 2017; McNamara, 2012), although there is not clear consensus among the literature (Starosta & Izydorczyk, 2020). Binge-watching has become far more common within the last decade, with the usage of the term booming in popular media between 2012 and 2016 (Steiner, 2018), during which it was named one of the most popular new words by the Oxford Dictionary (2013), and the Word of the Year by the Collins Dictionary (2015). The increasing prevalence of the term coincides with the advent of video on demand (VOD) viewing in the United States, popularized by providers like Netflix, Hulu, and Amazon in 2012-2013 (for a review, see Jenner, 2016). These entertainment companies began offering massive collections of shows, accessible to viewers at any time for a monthly fee. Consumers could choose to watch whole seasons of their favorite shows in just a few days. Taking advantage of this new feature, 62% of American respondents indicated that they had engaged in binge-watching (defined as two or more consecutive episodes; Harris Interactive, 2013), and this behavior has been increasing in popularity since (Starosta & Izydorczyk, 2020). For instance, in 2016, 73% of polled Americans indicated that they had binge-watched, with nearly 30% doing so on a weekly basis (Deloitte, 2017). Binge-watching behavior is especially rampant among younger generations, with nearly 90% of Millennials and Gen Z engaging in binge-watching at some point, and 40% binge-

watching weekly (Deloitte, 2017). Methods of watching have seemingly transitioned to be independent of the rigid scheduling that is a hallmark of traditional television. As a result of the convenience of VOD viewing, the opportunity for binging has never been better.

Binge-watching: Previous Research

Although research on binge-watching is relatively scarce, the aspects of the phenomenon that are often studied can be divided into three categories: motivations, personality traits, and risk factors or consequences (Starosta & Izydorczyk, 2020). Motivations for binge-watching are, unsurprisingly, complex, and multifaceted. Most research emphasizes the role of instant gratification, as individuals meet their hedonistic entertainment, engagement, and relaxation needs (Starosta & Izydorczyk, 2020). Other research indicates social connection, or yearning to feel like part of the fandom, as a motivator. Additionally, individuals may binge in response to a fear of missing out (Conlin, Billings, & Averset, 2016), to escape from reality (Starosta et al., 2019; Panda & Pandey, 2017), or to cope with negative emotions (Rubenking & Bracken, 2018; Flayelle et al., 2019 Castro et al., 2019). Several individual personality traits can compound these motivations, making binging especially likely. Specifically, people who demonstrate a lack of control (either as being impulsive or acting urgently) and who are sensation seeking (searching for arousing stimuli) are more likely to binge-watch (Flayelle et al., 2019; Riddle et al., 2017). Additionally, individuals with a higher need for cognition and greater fantasy empathy (ability to feel the emotion of fictional characters) seem to experience greater transportability into the narrative of show. This increased ability to immerse in the content has been a significant predictor of binge-watching frequency

(Anghelcev et al., 2021; Conlin, 2015). Ultimately, the profile for binge-watchers is not likely to be universal. Indeed, some research has categorized individuals who binge-watch into four groups (avid binge-watchers, recreational TV series viewers, unregulated binge-watchers, and regulated binge-watchers) that differ in type and intensity of motivations and personality characteristics (Flayelle et al., 2019).

But with binge-watching comes consequences. Primarily, previous research has homed in on symptoms of behavioral addiction that are associated with binging. Some of the common risks associated include a loss of self-control, a feeling of regret, and a tendency to neglect responsibilities (Starosta et al., 2019; Flayelle et al., 2019; Riddle et al., 2017). In addition to these, binge-watching is usually a sedentary behavior that is related to reduced physical activity and consumption of unhealthy food (Spruance et al., 2017). Frequent binging has also been associated with poorer sleep quality and duration (Exelmens & Van den Bulck, 2017).

Most risks that have been found to be associated with binge-watching take the form of negative behavioral outcomes. Little empirical research has been done investigating cognitive consequences of binge-watching, however. A popular notion is that 'TV rots your brain,' and the more of it you consume, the worse off you are. However, the exact nature of this 'rot' is not well specified. Previous research has found declines in executive function and processing speed associated with three or more hours of daily viewing (Hoang et al., 2015). However, these results were attributable to increased sedentary behavior and decreased physical activity that typically accompanies viewing, rather than the viewing itself. Other studies have attempted to control for sedentary behavior and found that increased television consumption is associated with a

decline in verbal memory (Fancourt & Steptoe, 2019). Some studies have found poorer short-term memory (Bakrania et al., 2018) and mild cognitive impairment (Wang et al., 2006) to be associated with binging and sedentary behavior. In contrast, other research has found no such association (Hamer & Stamatakis, 2014; Kesse-Guyot et al., 2012).

Current Research Questions

There are two problems here that justify a need for further research into the effects of binge-watching on cognitive functioning. First, the lack of consensus in the literature is a somewhat obvious indicator that there is more nuance to uncover. Second, even if the current research was more consistent, these studies are still limited by their methodology. They provide questionnaires that rely on self-reported data and can conclude, at best, significant associations between variables. Research that experimentally manipulates an individual's watching behavior would help to investigate causality. This presents an important, unexplored direction for binge-watching research, and is the primary goal for this study. Specifically, we aim to investigate how experimentally manipulating watching behavior affects memory for what was watched. If someone were to watch several episodes of a show back-to-back in one sitting (i.e., binge-watch), would they have worse memory for that show than if they watched one episode per week? Phrased another way, would those who had spaced out their watching have better memory for the show? To gain insight into these questions, literature on the spacing effect, a memory and learning phenomenon, is reviewed next. Information about typical spacing research will be presented, several prominent theoretical explanations will be identified, and those accounts will be applied to the binge-watching context to make predictions. Following that, research will be presented to address these questions, with

the undertaken study serving as a replication and extension of previous unpublished work by Fogler and colleagues (2017).

Literature Review: The Spacing Effect

The spacing effect is defined as superior memory for information that is presented repeatedly across spans of time relative to information presented in mass (sometimes referred to as 'cramming'). The benefits of memory can be broken down into four categories: acquisition, retention, relearning, and generalization. To best address the primary research question, retention is the dependent variable of interest in this paper. Typical research designs to assess the spacing effect consist of a few core components: 1) An initial presentation of to-be-learned material, 2) followed by a subsequent relearning of the material (i.e., a study session) and 3) a memory assessment for the material. There are two critical intervals in this setup. The first, a *lag* or *spacing gap,* refers to the length of time between the initial learning session and relearning session (this is sometimes also termed the interstudy interval or ISI for short). The second is the time between the final learning session and the memory assessment, termed the *delay* or *retention interval*.

The Spacing effect is a relatively robust phenomenon, occurring across a wide range of domains (see Dempster, 1988). It has been demonstrated with free recall (Thios & D'Agostino, 1976), recognition (Hintzman & Block, 1970), or paired-associate learning (Greeno, 1964). Additionally, the findings from laboratory experiments have translated well into real classroom contexts (see Hopkins et al., 2016; Dunlosky et al., 2014; Sobel, Cepeda, & Kapler, 2011). The benefits are emphasized heavily in the field of education, where teachers are encouraged to structure their courses to allow for spaced out practice of material and students are encouraged to utilize spaced retrieval in their

studying (Pashler, 2007). However, the spacing effect has not been directly studied as it relates to binge-watching VOD content. As such, it is necessary to examine the different theoretical accounts that attempt to explain the spacing effect, given that they may lead to different predictions when applied to this novel milieu.

Theoretical Accounts

The robustness of the spacing effect and the differential benefits across various categories make it difficult to synthesize and understand the underlying mechanisms operating during spaced practice. As a result, there is no overarching, widely accepted explanatory account that addresses the many caveats of research findings. Instead, there are several prominent competing theories that are not necessarily mutually exclusive. Namely, these theories revolve around contextual variability, deficient processing, and study retrieval.

Contextual Variability

According to theories of contextual (or encoding) variability, contextual elements present at the time of study are stored alongside the studied item. As such, the memory trace consists of the item itself and the corresponding contextual details. Spaced repetitions lead to a greater variety of contextual elements being encoded alongside the to-be-learned information than do massed repetitions (Melton, 1967, 1970; Madigan, 1969; Bower, 1972; Raajimakers, 2003; Maddox, 2016) because the repetition contexts are increasingly varied as a function of time. With more contextual elements being available to cue retrieval, there is a greater likelihood of overlap with contextual elements present at the time of the test, and spaced learners should then be better able to cue and recall memories after delays. As an example, Glenberg (1979) devised an experiment in

which participants studied related word pairs (e.g., blade-knife). In the constant condition, participants studied the same word pairs in each of two study sessions (i.e., they saw blade-knife each time). In the varied condition, the second presentation of the word pairs differed from the first. The cue word was replaced by a new word that was semantically related to the target. As a result, participants in this condition may see 'blade-knife' in the first study presentation, but 'spoon-knife' in the next. Participants in the varied condition demonstrated a greater proportion of target words recalled than participants in the constant condition. This finding provides evidence that varied encoding (through two different associations) contributed to better memory for target words.

Deficient Processing

The next group of theories suggests a declining ability to effectively encode material when it is massed together with little-to-no lag between presentations. These theories are referred to as deficient (or diminished) processing accounts. They ascribe poorer memory for massed material as being due to an impaired encoding process for subsequent information presented shortly after the initial information. Hintzman's (1974) inattention theory serves as a prime example of such an account, in which subsequently presented material is not properly encoded because it is poorly attended to. This inattention could be viewed as somewhat voluntary, with learners being less inclined to engage with the material (i.e., practice retrieval) in massed settings because it already seems familiar to them (Bahrick & Hall, 2005). The findings of Shaughnessy, Zimmerman, and Underwood (1972) support this claim. The researchers allowed the participants to control the pace at which to-be-learned items were presented by giving

them an option to proceed to the next item at their discretion. Participants spent significantly less time studying repetitions when they were massed together versus spaced apart. Participants also recalled a greater percentage of items when their repetitions were spaced out. Taken together, these results may indicate that individuals demonstrate greater willingness to attend to repetitions of material when it is spaced out rather than massed together. This inattention to massed repetitions could then be said to result in diminished processing.

Study-Retrieval

The final class of theories presented here are those emphasizing study-retrieval. From this perspective, the second presentation of material cues the participants to retrieve the memory trace of the first presentation. This retrieval facilitates later activation of the memory trace, and thus enhances future recall. Evidence for this approach can be seen in a series of experiments by Thios and D'Agostino (1976) in which participants studied short sentences written in active voice. All participants studied the sentences twice, but for the second presentation, were instructed to alter the sentence to be in passive voice and say it aloud (e.g., "the conductor boarded the express train" would be reformulated into "the express train was boarded by the conductor). However, one group of participants was given the full initial sentence to aid in their reformulation, and the other (the retrieval group) was given only the object phrase (i.e., "express train"). To complete the task, the latter group would have to rely on retrieving the trace of the first presentation, while the former group would not. Additionally, three different spacing intervals were added as a between subject variable. Afterwards, participants were given time for free recall of the object phrases. The proportion of correctly recalled object

phrases increased as a function of the spacing interval, but only for participants in the retrieval group. In other words, the spacing effect was present, but only in the group that retrieved the first presentation. This was early evidence of the benefit of retrieval practice, and there is a large body of literature regarding this phenomenon today (see Carpenter, 2011, for more information about retrieval practice in the testing effect).

Contextual variability, deficient processing, and study-retrieval are three foundational ideas for explanatory accounts of spacing. However, there are many combinations (see Greene, 1989), and variations of these ideas (for a consolidation account, see Lehmann et al., 2009; Smith & Scarf, 2017), as well as additional considerations (for the effect of retrieval difficulty, see Bjork & Bjork, 1992; Delaney, Verkoeijen, & Spirgel, 2010). While a complete picture of theories and mechanisms underlying the spacing effect is likely far more complex than presented here, a full review of the spacing literature is outside of the scope of this paper (for more in-depth reviews, see Cepeda et al., 2006; Benjamin & Tullis, 2010; Delaney et al., 2010).

Applications for Binge-Watching

Predictions by Theory

An attempt to integrate the spacing effect's leading theories into the area of binge-watching yields mixed results. Episodes of a show, when spaced out over longer intervals, should produce viewing sessions that are more contextually diverse than episodes watched in one massed sitting. The greater variety of contextual components would lead to a greater number of encoded stimuli associated with the target information. Similar to how Glenberg (1979) found that increasing the number of associations (blade, spoon) with target words (knife) resulted in greater memory for the targets, it could be

reasonably surmised that the greater number of contextual associates for content in spaced episodes would result in better memory for that content. However, it is unclear whether the general wealth of contextual information present in the episodes (in the form of embellishing details), regardless of whether they are viewed in spaced or massed conditions, would contribute to diminishing returns of additional contextual information provided by spacing out viewing. In other words, perhaps episodes viewed in mass still contain a plethora of contextual elements that aid in later retrieval, making the benefits of spacing go relatively unnoticed.

From a deficient processing perspective, predictions for spacing benefits in binged content may be similarly caveated. In the example mentioned earlier, Shaughnessy, Zimmerman, and Underwood (1972) demonstrated that participants choose to spend more time studying repeated information when it is spaced out instead of massed. It was suggested that this preference was evidence of inattention for massed repetitions. In the case of binge-watching, where episodes are not direct repetitions of information (discussed in more detail later), it is unknown whether participants will share a similar unwillingness to attend to massed presentations. In fact, a qualitative analysis of binge-watching motivations by Peterson (2016) revealed that many people report the opposite when they are binge-watching. Being absorbed in the show's content was a common theme that aided watchers in escaping from reality and encouraged them to keep watching. Some individuals reported that they were so attentive to their show that they became relatively oblivious to their surroundings. This sort of tunnel visioning seems to starkly contrast with theories of voluntary inattention. If deficient processing results in

poorer memory for binged content, it may be through involuntary means (see Hintzman, 1974).

From a study-retrieval perspective, spacing out episodes of a show would result in more instances of effortful retrieval of show content than watching in mass. This would occur because the presentation of subsequent episodes would, theoretically, cue some retrieval of previous episode content, whereas massed viewing would require less effortful retrieval. Because the content presented in subsequent episodes builds on previous episode content, rather than directly repeating it, it may even be argued that retrieval of previous episode content is *more* likely than in traditional spacing study sessions. In the example from Thios and D'Agostino (1976), participants in the retrieval group saw only partial repetitions of previously studied material, then had to construct a new (albeit related) sentence with that material. That is, they needed to retrieve the trace from the first presentation of the material to complete the task given them at the second presentation. Following this logic, information in subsequent episodes that builds upon information presented in the initial episode would similarly rely on retrieval of the first episode. However, it is unclear if this retrieval process occurs differentially between spaced viewing and massed viewing. It is likely that the retrieval of content from the initial episode would be more difficult as the gap between episodes increased (i.e., in spaced viewing). According to theories of retrieval difficulty, the advantage of retrieval practice is most apparent when the memory traces are difficult to retrieve (Bjork & Bjork, 1992; Delaney, Verkoeijen, & Spirgel, 2010). As such, theories of study-retrieval could be combined with theories of retrieval difficulty to predict benefits of spacing out episodes. Additionally, a systems consolidation approach that emphasizes the benefits of

time and sleep on consolidating memory traces in the hippocampus and cortex could be combined with study-retrieval theories to yield a similar prediction.

Additional Considerations

In addition to research revolving around explanatory theories of spacing, there are several previous studies that demonstrate necessary elements contributing to the hypotheses previously presented here. First, the benefits of spacing are not attenuated by the modality of the presentation of to-be-remembered material. The benefits of spaced learning have been demonstrated when to-be-learned material are presented only visually, only auditorily, or bimodally (auditory plus visual) (see Melton, 1970; Hintzman, Block, & Summers, 1973; Janiszewski, 2003). Additionally, the effect size is not demonstrably influenced by the use of pictorial material instead of verbal material (Janiszewski, 2003). As such, the bimodal pictorial presentation of material that occurs when watching episodes of a show is not expected to have any detrimental influence on the benefits of spaced viewing over binge-watching.

Outside of the presentation format, there are still some features of this study that make it atypical. Traditional spaced learning sessions consist of repeating identical content (for instance, the same 20 word pairs). For the purposes of this study, individual episodes in a series are being equated to traditional learning (or studying) sessions. However, the content presented in subsequent episodes of a series is not identical to the content presented in prior episodes. There are similar individual elements carried through the episodes (characters, settings, themes, etc.), but the stories that make up the larger picture build on previous episodes, rather than being identical to them. There are a few ways to approach this difference. A meta-analysis by Janiszewski, Noel, and Sawyer

(2003) found that the benefit of spacing when stimuli were presented in isolation (i.e., one word or picture at a time) was similar to when stimuli were embedded in contextual details (i.e., as part of a sentence, or piece of a larger image). Episodes from a show may then be viewed as numerous bits of to-be-learned target stimuli housed within a large amount of contextual details. As such, it may be feasible to expect that memory for explicitly repeated stimuli across episodes will be influenced by the distribution of episode viewing. Still, it is unclear whether that expectation could be held for information that is less explicitly repeated, and more built upon. In a guide for improving student learning, from the U.S. Department of Education, the authors state "In certain classes, important content is automatically reviewed as the learner progresses through the standard curriculum and [spacing learning over time] may be unnecessary in courses where this is the case" (Paschler et al., 2007, p. 5). The authors then give the example of students using single digit addition every day in second grade math classes. In these instances, in which to-be-learned skills and concepts are built upon by subsequent learning (as in mathematics), the researchers reported that additional distributed practice may not be needed. This seems to imply that the building process entails an automatic review of previously presented material. If this is the case, then the benefits of spaced distribution of each of those building sessions should be somewhat akin to benefits of spacing distribution of identical repeated material. If the difference is not relatively negligible, it seems unlikely that the U.S. Department of Education would suggest the latter be unnecessary when the former is already in place. Still, it is unclear how this distinction influences the benefits of spacing as it applies to this paper. As such, possible theoretical implications (mentioned above) will be considered, and analysis of all data

gathered from the following design will be compared with the data from only explicitly reoccurring items.

Predictions Summarized

The three prominent theories discussed have a wealth of supporting evidence and emphasize the robustness of the spacing effect across many domains and modalities. Binge-watching is the opposite of spaced learning, which could imply that there is a deficit in memory for binged content. Two of the theories discussed would predict such a deficit. First, spaced presentations of episodes provide more contextual variability than massed presentations. As such, contextual variability accounts would predict reduced or impaired memory for binged content. Next, study-retrieval accounts emphasize the benefit of retrieval practice in contributing to a spacing effect. Because subsequent episodes of a show consistently build on content in previous episodes, they are apt to cue retrieval of previous episode content. The less effortful retrieval occurring in the amassing of episodes would lead to reduced or impaired memory for binged content. However, massed presentations may not undergo deficient processing, which has been attributed to inattention in past studies, due to the engaging nature of the show. In other words, people who are binge-watching may be fully attentive to the show, thereby eliminating a mechanism through which deficient processing is reported to occur. When viewed from this lens, deficient processing accounts would predict no impairment in memory for binged content. Giving credence to the robustness of the spacing effect, however, it seems more likely that binge-watchers would have poorer memory for show content than spaced viewers. Yet, when this hypothesis was tested by Fogler and colleagues (2017), such an effect was not found. Participants who viewed episodes of a

show in mass had similar memory for the content as those who viewed them spaced out over several sessions.

Extending Fogler and Colleagues' Previous Work

Before diving into the theoretical implications that this leads to, there could be several reasons as to why Fogler and colleagues (2017) found no effect. First, their memory quiz consisted of multiple-choice questions, thus only assessing recognition memory. Typical spacing studies incorporate short answer questions that more directly target recall (either cued or free; see Cepeda et al., 2006). Although both question types would require some sort of retrieval to answer correctly (apart from guessing), recalling the information would be more difficult and effortful than simply being able to recognize it as familiar. With retrieval difficulty theories in mind (Bjork & Bjork, 1992), it could be argued that the benefits of spacing are then more likely to be seen in recall tasks. However, there have been instances where no discernible difference in effect size was found between recall and recognition tasks (Janiszewski, 2003). Still, it would be wise to incorporate recall questions that mirror more typical spacing methodology.

Fogler et al. (2017) additionally may have seen reduced spacing benefits due to the joint effects of their length of delay (retention interval) and spacing gap (ISI). Previous researchers (Cepeda et al., 2006; Cepeda et al., 2008) have identified an inverted-U shaped relationship between spacing gap and proportion correctly recalled. In essence, increasing the spacing between learning sessions (spacing gap) improves retention up to a point. When spacing gaps are lengthened beyond this optimal point, retention begins to decrease. However, the optimal point differs according to the length of the delay, with longer delays needing longer spacing gaps to achieve optimal retention.

For a delay of seven days, as seen in the work of Fogler and colleagues, a spacing gap of approximately one day maximizes retention (Cepeda et al., 2006). Fogler and colleagues used a spacing gap of seven days however, consistent with traditional weekly television schedules. As a result, the benefit of spacing was likely attenuated to some degree. In order to more fully investigate whether spacing effects exist in binge-watching, it is necessary to try and create conditions that would maximize the benefits of spacing. As such, a reasonable next step is to manipulate either the delay or spacing gap in a way that better optimizes spacing effects. If the spacing gap is left at seven days to mirror traditional television installments, then a delay near thirty days would optimize spacing advantages (Cepeda et al., 2008). Beyond this, the nature of this research is largely exploratory. It may be the case that binge-watchers do not experience greater deficits in memory (relative to spaced viewers) with shorter delays, but these deficits become pronounced with time. As such, perhaps Fogler et al. (2017) found no difference between spaced and massed conditions because their one-week delay was not long enough for an effect to emerge. Adding a longer delay would thus serve an essential investigative function.

The Current Study

In weighing the robustness of the spacing effect, Hintzman (1974) states "the investigator who fails to obtain the effect would probably be wise to suspect sampling error, ceiling effects, or a flaw in the experimental design" (p. 79). In accordance with similar reservations, it may be equally wise to replicate Fogler et al. (2017) before seriously interpreting their findings. As such, this study serves as a replication with several extensions. Individuals will watch episodes of the Netflix original series, House

of Cards, either in one three-hour session, or across three separate sessions over three weeks. Following, each group will be tested on their memory for the episodes using a scale developed by Fogler and colleagues (2017) and modified for the purpose of this study. As an extension of their work, this study further aims to compare both recognition and recall memory across the viewing conditions as well as assess whether the spacing effect is present with a longer retention delay. These goals are accomplished by the inclusion of short answer questions on the memory assessment and the addition of a four-week test delay group.

Hypotheses

It is still hypothesized that individuals who watch episodes spaced out over several sessions will have better memory for show content than individuals who watch the same episodes in one massed session. Additionally, it is predicted that participants with a four-week delay between viewing and testing will demonstrate poorer memory for the episodes, when compared with the one-week delay group, given the decrement in memory as time passes (Ebbinghaus, 1885). Finally, the presence of an interaction between viewing condition and delay group is anticipated, as it is hypothesized that a spacing effect will be stronger in the longer delay group (because of the joint effects of retention interval and spacing gap on optimal retention; Cepeda et al., 2006; Cepeda et al., 2008).

Method

Participants

A total of 120 participants were recruited from the undergraduate population of a moderately sized public university located in the southeastern United States. Using

statistical software GPower (Erdfelder et al., 1996), a priori power analyses showed that a sample of at least 128 participants (32 per group) would be necessary to obtain adequate power (1-β = .80) to detect an effect (discussed more in the limitations section). Recruitment occurred via an online participant pool, alongside other options, and participation was rewarded with course credit. Sixteen participants dropped out of the study after consenting to participate. Ten participants were excluded from analyses due to reporting that they had seen the episodes used in this study (Season 1, episodes 1-3) prior to participation. Of these participants, two watched the episodes 1 to 6 months prior to participation, three watched them 6 to 12 months prior, and five saw them a year or more before participation. Finally, data from two participants was excluded from analyses because the participants did not indicate whether they had seen the show before.

Data from 92 participants (70.7% female, 29.3% male) was therefore used for analyses. They were majority White (72.8%), followed by Black or African American (13%), then by Hispanic or Latino (6.5%), and Asian (5.4%). A small fraction of participants identified as multi-ethnic (1.1%) or 'other' (1.1%). Participant ages ranged from 18 to 33 years (M = 19.2 years, SD = 2 years). Most participants reported being freshmen (68.5%), followed by sophomores (14.1%), juniors (12%), and seniors (5.4%).

Materials

The three component materials utilized within this study are: a streaming service that allows easy access to video content, an assessment of participant memory for that content, and a demographics measure that includes questions about participant binge-watching behavior. Netflix was chosen to satisfy the first need, and the Binge-Watching

Episode Assessment (BWEA) was created to meet the second. Demographics and additional binge-watching questions were attached to the end of the BWEA.

Netflix's 'House of Cards'

The streaming service, Netflix, was utilized to project video content to participants. Specifically, the first three episodes of season one of Netflix's original series, House of Cards, were shown. House of Cards, a political thriller, was selected for this research because it was deemed less likely to have been previously watched by an undergraduate population, while still being well-reviewed by critics. Additionally, some research shows that dramas are the most binged genre (Deloitte, 2018). Each of the first three episodes is approximately 50 minutes in length, with total watch time equaling 156 minutes. Additionally, the episodes did not contain a recap or summary of previous episode content, as is seen in some shows.

The Binge-Watching Episode Assessment

In order to assess memory for the episodes, a cognitive scale was created by Fogler et al. (2017) using surveying software, Qualtrics, and modified for the purposes of this study. The modified Binge-Watching Episode Assessment (BWEA; See Appendix A) consists of 40 questions that all pertain to events that explicitly occurred within the first three episodes of House of Cards or could be inferred from them. Questions 1 through 30 are multiple choice, and the last 10 questions are counterbalanced across two versions of the assessment (with 5 multiple choice and 5 short answer on each) so that recognition and recall can be properly compared across the same question stem. Each question was weighted equally, and a higher score on the measure indicates better memory for events in the episodes. Following the cognitive assessment, the BWEA has

three items assessing whether participants have seen these episodes before, several qualitative items in which participants define 'binge-watching,' and demographic questions.

Despite being a new, researcher generated instrument, initial analysis of the psychometric properties of the BWEA (based on a sample of 150 test-takers) revealed that it has moderate internal consistency, $\alpha = .70$. Additionally, the standard error of measurement, or the average deviation of an observed score from a true score, was relatively low (2.11), indicating high precision. A table of further analyses at the item level is attached in Appendix B, and a discussion of the scale validity can be viewed in Austin and Shapovalov (2020). It is important to note that these psychometric analyses were conducted on only the first 30 questions of the BWEA, due to sample size limitations for the last 10 items (see Austin & Shapovalov, 2020, for more detail). As such, a psychometric analysis of the full scale is recommended as a future goal for research in this area.

Demographic and Binge-Watching Behavior Questions

Basic demographic questions were included to garner information on participant race, gender, age, and year in school. Following these were several critical questions about participant's previous exposure to the show, including "Have you watched House of Cards at any point prior to this study?" and "Have you seen these episodes at any point prior to this study?" in addition to "How long ago did you watch these episodes?" Additional questions assess participant definitions of binge-watching, as well as their individual motivations for engaging in binge-watching, and frequency with which they binge-watch.

Design

A 2x2 between-subjects design was implemented. The independent variables (Viewing, Delay) each have two levels (Spaced, Massed; 1-week, 4-week) that result in four independent groups. Participants were each randomly assigned to one of these four groups. The spaced groups viewed episodes once a week for three weeks, while the massed groups viewed the same episodes back-to-back in a single three-hour session. Neither group was given a summary or review of previous episode content. The 1-week delay groups were administered the BWEA one week following the presentation of the last episode, while the 4-week delay groups were administered the BWEA four weeks after the last episode. The dependent variable, memory, is operationalized as the total score (0-35) on the recognition questions from the BWEA. For secondary analyses, memory is also operationalized as the total score (0-5) on the recall questions of the BWEA as well as the score on recurring content multiple choice questions only (0-7). Finally, to hold the delay between viewing and time of test constant across groups, memory is operationalized as total score on questions from the last episode only (0-5).

Procedure

Upon entry into the room, participants were directed to sign in with their name and university identification so that they could be awarded credit for their attendance. Afterwards, informed consent forms were read and signed. At this point, individuals were given instructions requesting they put away their cell phones and reminding them to pay attention to the episodes presented because they will be tested on their memory for this content (see Appendix C for standardized instruction and protocol). Following these

instructions, the door was closed, and the lights dimmed. The researcher(s) then began presenting episode 1 of House of Cards.

Upon completion of the first episode, procedures differ depending upon which viewing condition the participant was randomly assigned to. Participants in the spaced group were given brief instruction and dismissed until the next week. They came back approximately one week later (after viewing episode 1) to view episode 2 according to a similar procedure as used for episode 1. Following this session, they met again one week later (after viewing episode 2) to view House of Cards episode 3. The massed group, on the other hand, took a 5-minute break following the conclusion of episode 1, prior to the presentation of House of Cards episode 2 (massed). The massed group then took another 5-minute break following episode 2, before they begin episode 3, after which they were given brief instructions and dismissed. As a result, the massed group viewed all three episodes in a single, 3-hour session, while the spaced group viewed one episode per session over a period of three weeks. After each group viewed episodes 1, 2, and 3 of House of cards, they concluded the viewing phase of the experiment. Participants were then instructed to avoid viewing any episodes of House of Cards (outside of the experimental sessions) until after they have taken the BWEA.

After the viewing phase, participants were administered the BWEA. Some participants took the assessment one week after completing the first phase, while others took it four weeks after completing the first phase. Due to complications arising from the coronavirus pandemic shortly after the start of data collection, some participants ($n = 32$) completed the BWEA in person, while others completed it online ($n = 60$). In-person participants met in a computer lab at their designated time to complete the BWEA. Upon

arriving, the assessment was preloaded onto an adequate number of computers in the room, and already displaying the first page of the test. Before beginning, participants were given brief instructions to discourage cheating and help them navigate the online quiz format (See Appendix C). They were also informed that they have 45 minutes to complete the assessment, although participants typically finish in about 20 minutes. The participants taking the BWEA online were emailed the same instructions along with a link to the assessment. Upon completion of the BWEA, individuals were thanked for their participation and dismissed, concluding their role in the study.

Results

Recognition Memory

Descriptive statistics for recognition memory scores separated by delay and viewing condition can be seen in Table 1.

Table 1

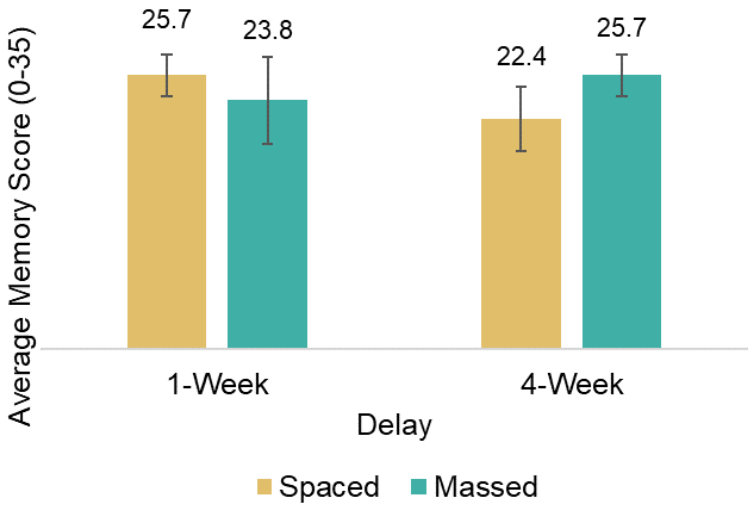
Descriptive Statistics for BWEA Multiple Choice Questions (0-35)

Delay	Viewing condition	N	*M*	*SD*	Skewness	Kurtosis
One-Week	Spaced	29	25.7	4.2	-0.79	-0.05
	Massed	20	23.8	7.1	-1.12	0.55
Four-Week	Spaced	20	22.4	5.2	0.18	-1.09
	Massed	23	25.7	3.6	-0.29	0.26

At a one-week delay, spaced viewers outperformed massed viewers. At a four-week delay, however, the opposite was true, with massed viewers outperforming spaced viewers. Figure 1 illustrates these relationships visually. Relative to the other groups, the massed four-week delay group averaged higher scores than was anticipated given previous literature, roughly equaling that of the spaced one-week delay group.

Figure 1

Average Memory Score Separated by Viewing Condition and Delay



Note. Error bars represent 95% confidence intervals.

To investigate group differences, a 2 (viewing: spaced vs. massed) x 2 (delay: 1-week vs. 4-weeks) between-subjects factorial analysis of variance (ANOVA) was conducted, examining the effect of viewing condition and delay on memory. The assumption of homogeneity of variance was met, Brown-Forsythe $F(3, 88) = 2.409$, $p = .072$. Normality was violated in two groups (see Table D1). However, given our relatively balanced design with moderately large group sizes, and consistent direction of skew for the distribution of scores in each group, our ANOVA is likely robust to this violation. All other assumptions were met, and the Type I error rate is set to be .05.

The pattern of differences in memory scores between spaced and massed viewers depended upon the length of the delay, $F(1, 88) = 6.103$, $p = .015$, $\eta_p^2 = .065$, 1-$\beta = .69$. There was no main effect of viewing condition, $F(1, 88) = 0.432$, $p = .513$, $\eta_p^2 = .005$, 1-$\beta = .10$, or delay, $F(1, 88) = 0.425$, $p = .516$, $\eta_p^2 = .005$, 1-$\beta = .10$. To follow up on the

significant interaction, simple comparisons were done at each level of delay. At one week, the difference between spaced and massed viewers was not significant, $t(47) = 1.195$, $p = .238$, $d = 0.35$, $1\text{-}\beta = 0.22$. At a delay of four weeks however, the massed group ($M = 25.7$, $SD = 3.6$) scored significantly higher than the spaced group ($M = 22.4$, $SD = 5.2$), $t(41) = 2.481$, $p = .017$, $d = 0.76$, $1\text{-}\beta = 0.68$. The effect size indicates that the difference between groups amounts to 76% of a standard deviation. Thought about in another way, the massed viewers scored roughly 9 points higher than the spaced viewers on the percent correct metric (73%; 64%).

Explicitly Recurring Content

To better align with spacing literature, secondary analyses were performed to examine the effects of viewing condition and delay on explicitly recurring content only. Memory for recurring content was measured using seven multiple choice questions pulled from the BWEA. Descriptive statistics separated by viewing condition and delay can be found in Table 2. The group means are relatively clustered, with the spaced one-week delay group averaging slightly higher memory scores than the others. There is a similar pattern of scores as seen above from analyses with all recognition questions. At one-week delay, the spaced viewers scored above massed viewers, but at four-weeks, the pattern is flipped. However, the difference between massed and spaced at four-weeks is very small and likely negligible in this case.

Table 2

Descriptive Statistics for BWEA Recurring Content Questions (0-7)

Delay	Viewing Condition	N	*M*	*SD*	Skewness	Kurtosis
One-Week	Spaced	29	5.5	1.4	-0.54	-0.84
	Massed	20	4.8	1.6	-0.68	0.15
Four-Week	Spaced	20	4.9	1.2	-0.11	-0.96
	Massed	23	5.0	1.3	-0.40	-0.05

A 2 (viewing: spaced vs. massed) x 2 (delay: 1-week vs. 4-weeks) between-subjects ANOVA was utilized. Normality was violated in the spaced one-week delay group but satisfied in each of the other groups (see Table D2). Homogeneity of variance was satisfied, Brown-Forsythe $F(3,88) = 0.361$, $p = .782$. Results from the ANOVA indicated no significant interaction of viewing condition and delay, $F(1,88) = 1.865$, $p = .175$, $\eta_p^2 = .021$. Additionally, there was no main effect of viewing condition, $F(1,88) = 0.727$, $p = .396$, $\eta_p^2 = .008$, or of delay, $F(1,88) = 0.464$, $p = .497$, $\eta_p^2 = .005$.

Recall Memory

Further analyses were performed to examine the effects of viewing condition and delay on recall memory, as measured by five short answer questions. Descriptive statistics can be found in Table 3. The group means were similar across all groups, with the massed four-week delay group averaging slightly lower scores than the others. However, scores were markedly low across all groups, likely due to the inclusion of questions asking about proper names (of characters, and businesses) being especially difficult to recall. For instance, questions 37 and 39 (See Appendix A) saw few participants provide correct answers (11% and 17% respectively).

Table 3

Descriptive Statistics for BWEA Short Answer Questions (0-5)

Delay	Viewing Condition	N	*M*	*SD*	Skewness	Kurtosis
One-Week	Spaced	29	2.6	1.2	0.90	0.36
	Massed	20	2.5	1.4	0.21	-0.38
Four-Week	Spaced	20	2.4	1.3	-0.23	0.08
	Massed	23	2.1	1.2	-0.27	-0.55

A 2 (viewing: spaced vs. massed) x 2 (delay: 1-week vs. 4-weeks) between-subjects ANOVA was again utilized. The assumption of normality was violated by two groups (see Table D3), while homogeneity of variance was satisfied, Brown-Forsythe $F(1,88) = 0.283$, $p = .838$. Results indicate no significant interaction of viewing condition and delay, $F(1,88) = 0.064$, $p = .801$, $\eta_p^2 = .001$. Additionally, there was no main effect of viewing condition, $F(1,88) = 0.337$, $p = .563$, $\eta_p^2 = .004$, and no main effect of delay, $F(1,88) = 1.323$, $p = .253$, $\eta_p^2 = .015$. Given that there were already so few short answer questions, and a portion of them were incredibly difficult, interpreting the results of these statistical tests should be done cautiously. It is unlikely, for instance, that these questions are adequate representations of recall memory for the show content. As such, it is expressly emphasized that these findings are not necessarily indicative of an absence of the spacing effect in recall memory in this scenario. Ergo, these results should be interpreted as exploratory in nature.

Episode 3 Content

Finally, to hold the delay between viewing and time of test constant across groups, questions from only the last episode were analyzed. Descriptive statistics can be found below in Table 4. There were only five questions specific to episode three;

however, the pattern of memory scores was similar to that seen when all recognition questions were analyzed. Spaced viewers performed better than massed at a one-week delay, whereas massed viewers outperformed spaced at a four-week delay.

Table 4

Descriptive Statistics for BWEA Episode 3 Questions (0-5)

Delay	Viewing condition	N	*M*	*SD*	Skewness	Kurtosis
One-Week	Spaced	29	4.5	0.9	-1.67	1.99
	Massed	20	3.9	1.1	-1.11	0.83
Four-Week	Spaced	20	3.1	1.2	-0.11	-0.84
	Massed	23	4.1	0.8	-0.66	0.02

A 2 (viewing: spaced vs. massed) x 2 (delay: 1-week vs. 4-weeks) between-subjects ANOVA was used to investigate group differences in episode 3 memory scores. The assumption of normality was violated by three groups (see Table D4), while homogeneity of variance was satisfied, Brown-Forsythe $F(3,88) = 1.170$, $p = .326$. A significant interaction was present, $F(1,88) = 14.611$, $p = <.001$, $\eta_p^2 = .142$, such that the pattern of differences in episode 3 memory between spaced and massed viewers depended upon the length of the delay. There was a main effect of delay, $F(1,88) = 7.369$, $p = .008$, $\eta_p^2 = .077$, indicating that when averaging across viewing conditions, memory was better at one-week ($M = 4.2$, $SD = 1.0$) than four-weeks ($M = 3.6$, $SD = 1.1$). There was no main effect of viewing condition, $F(1,88) = 1.052$, $p = .308$, $\eta_p^2 = .012$. To follow up on the significant interaction, simple comparisons were done at each level of delay. At one week, spaced viewers ($M = 4.5$, $SD = 0.9$) had significantly higher scores than massed viewers ($M = 3.9$, $SD = 1.1$), $t(47) = 2.043$, $p = .047$, $d = 0.59$. The effect size indicates that the difference between groups amounts to 59% of a standard deviation. Thought about in the percent correct metric, spaced viewers scored 12 percentage points

higher than their massed counterparts (89%; 77%). At a delay of four weeks, this pattern was reversed, with massed viewers ($M = 4.1$, $SD = 0.8$) scoring significantly higher than spaced viewers ($M = 3.1$, $SD = 1.2$), $t(41) = 3.320$, $p = .002$, $d = 1.02$. The difference between groups was slightly greater than one standard deviation. In terms of percent correct, massed viewers scored about 20 percentage points higher than their spaced counterparts (82%; 61%).

Discussion

The primary goal of the current research was to investigate the effect of binge-watching on memory. Specifically, methodology roughly mirroring that typically seen in spacing effect research was used to assess whether a spaced advantage (or massed disadvantage) exists when individuals consume subsequent episodes of a character-driven drama. Participants were assigned to view episodes of a drama in either a spaced or massed viewing condition, with either a one or four-week delay before they were tested on their memory for show content. Recognition memory, assessed by 35 multiple choice questions, was the primary dependent variable. Secondary analyses were done using seven multiple choice questions covering recurring content only, as well as with five short answer questions. Below, the substantive findings are summarized, and a possible explanatory account is offered. Afterwards, limitations and future directions for research are discussed.

Summary of Findings

It was hypothesized that individuals who watched episodes of a show spaced out over several sessions would have better memory for episode content than individuals who watched the same episodes in one massed session. In other words, a classic spacing effect

would be demonstrated, resulting in poorer memory among binge-watchers. However, the findings from this research offer little to support such a prediction. At a one-week delay, spaced viewers did have slightly better recognition memory than massed viewers. The magnitude of this difference, however, was small compared to the variability in scores. As a result, the two groups did not differ significantly, replicating the findings of Fogler and colleagues (2017). This finding was mirrored when explicitly recurring questions were used as the dependent variable. When only episode 3 questions were analyzed, a similar pattern was observed, with spaced viewers having better memory than massed at a delay of one week. This difference, however, was significant with a moderate effect size, and may indicate that a classic spacing effect is present at one week for all content, but is weakened by the confounding of condition and delay that is inherent to spacing methodology.[1] If this were the case, though, it would be expected that at a four week delay, analyzing only episode 3 questions would result in a similar finding. That is, spaced viewers would perform better, relative to their massed counterparts, and the advantage that massed viewers had at four weeks when all recognition questions were analyzed would be attenuated (or erased). This was not the case, however, as analyzing episode 3 content resulted in a stronger massed advantage at four weeks, as is discussed in the next section.

The four-week delay group was added as an extension of Fogler and colleagues' (2017) previous work to investigate whether a spacing effect would occur (or persist) with a longer delay. Given previous research on the joint effects of spacing gap and delay

[1] Wherein even at a delay of one-week, the spaced group is being tested on episode 1 content three weeks after it is presented, whereas the massed group has a true one-week delay for episode 1 content. Both groups, however, would have true one-week delays for episode 3 content.

(Cepeda et al., 2006), it was expected that the benefits of spacing would be stronger in the four-week group. The interaction that occurred was thus starkly different from what was anticipated. At a four-week delay, there was no benefit of spacing. Beyond that, the pattern was unexpectedly reversed, with massed viewers having better memory than their spaced counterparts. When all recognition questions were analyzed, the magnitude of this difference was meaningful, nearing 76% of a standard deviation, and 9 percentage points in the percent correct metric. Bringing this into an education context, this could mean a difference of a full letter grade (e.g., going from a 'D' to a 'C'). Further, when only episode 3 questions were analyzed, massed viewers had an even larger memory advantage, with a mean that was one standard deviation higher than that for spaced viewers. The difference was 21 percentage points in the percent correct metric, equating to a difference of two full letter grades (e.g., from a 'D' to a 'B').

Because of the ubiquity of the spacing effect, finding instances in which there is no benefit to spacing is quite valuable to further understanding of this memory and learning phenomenon. Our findings suggest that one such instance may occur when successive episodes of a character-driven drama are watched once a week versus being binged in one sitting. Not only does there seem to be no benefit of spacing on memory one week after the episodes, but there may be a benefit of binge-watching on memory at four-weeks. Provided that the finding here does reflect a genuine avenue through which spacing does not benefit memory, an important next step is to uncover why. Following, one possible explanation is introduced.

Attention as a Mediator

Deficient processing accounts emphasize the role of inattention as a primary mechanism through which incomplete, or impaired encoding occurs in massed presentations of material. Proponents of this account argue that if someone is presented material in mass then they will be less able to attend to it, and ergo less able to encode and retrieve it. However, if inattention is removed from the equation, and massed learners are instead as attentive as their spaced counterparts, it stands to reason that deficient processing should not occur. Therefore, spaced learners would have no memory advantage over massed learners. This could explain the lack of a meaningful spacing effect that was evident at a one-week delay. Perhaps the narrative format of episodes of a show (with connecting characters, themes, motifs in a temporal timeline) could counter any decrement in attention associated with massed viewing, maintaining comparable levels of attention between massed and spaced viewers. Furthermore, if the massed learners were *more* attentive than their spaced counterparts, could this also explain why massed learners had better memory than spaced learners (as was seen at four-week delay)? Recall that in previous research involving qualitative reports of college students who binge-watched (Peterson, 2016), many indicated a hyperattentive focus on show content to the extent that they became oblivious to their immediate surroundings. Previous research has defined this phenomenon as 'narrative transportation' (Gerrig, 1993; Green & Brock, 2000). Conlin (2015) states that transportation should increase "if a person has the ability to remain fully immersed in a narrative world, and the storyline is not interrupted" (p. 32). It could then be reasoned that longer gaps between episodes of a show would constitute an interruption in the storyline that hinders immersion and

transportation, while the lack of such gaps in binge-watching maintains story continuity and fosters attention and engagement. To this point, Conlin (2015) states:

> "Through binge-watching, the viewing experience of the audience is no longer interrupted at the end of every episode, allowing viewers to allot more of their cognitive resources to concentrating on the story and engrossing their emotions and thoughts on what is happening in the world of the TV show" (p. 19).

As such, transportation theory would predict binge-watchers to be more attentive of the same content than spaced viewers (with a notable exception for types of low attentiveness shows, Pittman & Steiner, 2019; discussed later in future directions). As a result, binge-watchers may have better encoded the show content, and thus have better memory for it at test. To this point, it should be made clear that we are not suggesting that individuals should choose to binge-watch for this advantage. We are, however, suggesting that enriched encoding as a result of binge-watching a character driven drama is an important finding for memory researchers and laypeople alike. Knowledge of the factors and conditions that contribute to this enriched encoding during massed presentations of material could have considerable implications, from informing theory surrounding the spacing effect, to practical insight regarding optimizing learning under time constraints.

Although the preceding discussion addresses the point of why binge-watchers may have equivalent or better memory for content than spaced viewers, through means of increased attention, it does not address why this may depend upon the length of the delay. It is presently unclear as to why there was a memory advantage to binge-watching at a four-week delay, but not at a delay of one-week. However, those results may be partially explained by pre-existing group differences. Specifically, it is possible that the massed

four-week group had an increased presence of *fantasy empathy* or *transportability*, both of which are personality traits that can be used to predict the likelihood of narrative transportation. Simultaneously, perhaps the massed one-week delay group was low on these traits. Scales to measure these traits were not included in this research, so this point is purely speculative. Further, through the use of random assignment into groups, this explanation is unlikely. However, including fantasy empathy or transportability scales to check that random assignment effectively controlled for pre-existing differences in these traits may be wise. Additionally, it would enable researchers to examine their role as potential moderators.

Limitations

There are several limitations with this study to consider. To begin, the study was underpowered. In order to have adequate power ($1\text{-}\beta = .80$) to detect an effect, a sample size of 128 participants (32 per group) was necessary (calculated in GPower; see Erdfelder et al., 1996). Given the sudden onset of the global coronavirus pandemic during data collection, participant recruitment was more difficult than anticipated. With only 92 participants total, our observed power to detect an interaction effect in the omnibus ANOVA was .69, while the power to detect each main effect was .10. Because a significant interaction was detected, the impact of the small sample size would be felt only in the failure to detect any main effects. However, given the small differences in means between each viewing condition (averaged over delay) and each delay (averaged over viewing condition), there likely was no main effect to detect. As such, the results of the primary omnibus test were not likely to have been negatively influenced by the lack of power. When the follow-up tests were run, however, the difference between memory

scores for massed and spaced viewers at a one-week delay was not negligible (See Table 1). However, a significant difference was not detected by the t-test, possibly because the observed power was very low (1-β = .22). As such, it is feasible the findings at a one-week delay would represent a typical spacing benefit if an adequate sample size were used.

In addition to this threat to statistical validity, there was an unanticipated confound caused by the coronavirus pandemic that threatens the study's internal validity. Individuals that participated after the pandemic took the memory assessment online, whereas those who participated prior to the pandemic took it in-person, in a campus computer lab. As a result, everyone in the spaced one-week delay group (n = 29), and some in the massed one-week delay group (n = 3) took the quiz in-person, whereas the remaining participants (n = 60) took the quiz online. This presents a difficulty in discerning what degree of influence in-person testing had on the scores from the spaced one-week delay group, as they had no online test-taking counterparts. While there are definitive differences between these two conditions, it is unclear the route through which the scores would be influenced. If, for instance, the in-person test-takers were administered the test in the same room as they watched the episodes, then we may expect their scores to be higher given the benefit of encoding specificity (see Tulving & Thomson, 1973). However, the test-takers were administered the test in a computer lab, separate from where they viewed any episodes, so this is not a concern. Differential effort or amount of cheating (i.e., looking up correct answers) could be possible between in-person and online testing, but it is not considered likely in this instance given the low stakes of the assessment and the standardized instructions. Finally, other spacing research

consisted of participants take the memory test either online or in-person, and found no differences (See Bell et al., 2014). As such, this confound exists within the study, but there's little reason to believe it is responsible for meaningful differences.

Future Directions

As was previously mentioned, future research surrounding memory for binge-watched content would benefit from including scales that measure fantasy empathy (see Davis, 1980) and transportability (see Dal Cin et al., 2004). These personality traits predict narrative transportation, and participants higher in these traits may be more attentive to the show content. Examining spaced and massed groups for pre-existing differences in these traits would thus enable researchers to investigate them as possible moderating variables in the relationship between viewing condition and memory for what was viewed.

Additionally, it was theorized that binge-watchers may be more attentive to the show content (fantasy empathy and transportability being constant), given the lack of story interruption between episodes preserving a sense of narrative transportation. This hypothesis, in line with deficient processing accounts, would thus suggest attention as a mediator in the relationship between binge-watching and memory. One possibility for examining this further would involve utilizing different types of shows. Steiner and Xu (2018) created a viewer attentiveness spectrum (VAS) that organizes shows by the amount of attention required to obtain certain gratifications. Dramas with complex plots (like House of Cards) are considered high VAS (HVAS), while sitcoms, game shows, or reality television are low VAS (LVAS). To obtain equivalent gratification, it is therefore necessary for watchers to be more attentive to dramas than game shows, for example. If

show type was manipulated in future research, then spaced and massed viewers could be compared on HVAS and LVAS shows. If attention is responsible for mediating the relationship between binge-watching and memory for a drama, binge-watchers would be expected to have worse memory for a game show than a drama. It is unclear if the spaced viewers would see a similar decline (resulting in no spacing effect), or no decline (demonstrating a spacing effect). Unlike dramas, game shows do not have a linear narrative with a complex plot. They may present a situation more akin to traditional spacing research, in which rather disparate facts are learned and relearned. As such, it may be expected that a traditional spacing effect would occur. Regardless, investigating how the findings of this study may or may not generalize to a variety of show content is a useful next step for furthering interpretation of these results, and may provide greater insight for theoretical implications.

Conclusion

The current study investigated how memory for the first three episodes of a character driven drama is affected by the method in which it is watched. The findings suggest that viewing one episode a week (spaced) presents little-to-no advantage over viewing each episode rapidly in a single three-hour session (massed). Further, massed viewers had better memory than their spaced counterparts at a delay of four-weeks. These results align with predictions from a deficient processing account, which emphasizes the role of attention as a mediating factor. It is therefore possible that binge-watchers do not experience memory deficits typically associated with massed presentation of information because they are better able to attend to it. Further, they may benefit from enriched

encoding as a result of hyperattention. This hyperattention may be fostered by the uninterrupted immersion into a narrative that binge-watching provides.

Appendix A

Binge-Watching Episode Assessment Cognitive Items

Please answer the following questions to the best of your ability. Please click the arrow to begin.

1. What are Frank Underwood's motives behind backing President-Elect Garrett Walker?
    a. Underwood believes Walker can change the country for the better
    b. Walker is a pawn in Underwood's plan to take control of the West Wing
    c. To move up the ladder and gain a new position and power
    d. To spite all those in Congress who do not support him
2. Why does Claire want two more tickets for the Jefferson Ball for the Holburns?
    a. Claire is trying to impress her new clients at the Clean Water Initiative
    b. Claire needs more donation money for the Clean Water Initiative
    c. Claire is trying to reconnect with her old friends
    d. Claire wants to impress them so they will help Frank with the education bill
3. Why does Underwood take the blame for the education bill draft being released to the press when talking to Donald Blythe?
    a. To make Blythe feel bad so Blythe will confront Linda about the far left draft
    b. To push Blythe to accept the blame and responsibility for the release of the draft
    c. To put Blythe in an uncomfortable situation so Blythe would draft up a new copy of the bill for Underwood
    d. To make Blythe feel bad so Blythe would resign from office
4. Why did the sermon that Frank and Claire went to relate to Frank's life?
    a. The sermon talked about marital struggles that had parallels to Frank and Claire's most recent argument
    b. The sermon talked about the most recent legislation that Frank was working on
    c. The sermon talked about faith, which Frank was lacking at the moment
    d. The sermon talked about rising when you are defeated, which is what Frank intends to do after not being nominated for Secretary of State
5. What did Zoe Barnes want from Frank Underwood?
    a. For Frank to be her mentor
    b. To get inside information for articles
    c. For Frank to divorce his wife
    d. To have a discussion on their shared interests
6. How did Frank Underwood get the education bill to Zoe Barnes even though it went through the shredder?
    a. Frank's colleague got it out of the trash

   b. The bill that Frank put in the shredder was not the education bill
   c. Frank had a second copy of the bill made so he could give it to Zoe
   d. The bill that Frank gave Zoe was fake
7. What is Frank's motive behind giving Zoe Barnes the editorial on the Camp Accords?
   a. To get Zoe's name in the press so she would become a credible source
   b. To link Garrett Walker to the editorial to undermine his presidency
   c. To link Michael Kern to the editorial in an attempt to remove him as a nomination for Secretary of State
   d. To bring down the newspaper that published the article so they lose their credibility
8. Why was what Frank Underwood said to the screaming man significant?
   a. Frank saying this to the man will be caught on camera and cause problems for Frank's political career
   b. It suggests that Frank is a heartless person
   c. The screaming man used to be Frank's boss, and this was Frank's way of belittling him just like the man used to belittle Frank
   d. It is symbolic of the fact that people who are not in a position of power do not have an influential voice
9. Why is Peter Russo sent to speak with Roy Kapeniak?
   a. Peter wants Roy to make a statement that Kern wrote the college article
   b. To get Roy to talk about what really happened in Gaza
   c. Peter wants Roy to make a statement supporting Frank Underwood as Secretary of State nominee
   d. To get Roy to speak about his previous relationship with Zoe Barnes
10. Why did Frank Underwood tell Zoe Barnes to write that the Walker administration will nominate Catherine Durant for Secretary of State?
   a. To stir up trouble, which is what Frank Underwood loves to do
   b. To punish the Walker administration with bad press, which is Frank's way of getting back at the administration for not picking him for Secretary of State
   c. So the media will spread it around, and the so the Waler administration will eventually pick Durant
   d. So the media will be misled, and Zoe Barnes will lose credibility
11. What realization does Zoe Barnes come to when Frank tells her to write a story about Senator Catherine Durant's nomination as Secretary of State in place of Michael Kern?
   a. That Underwood is using her to release negative stories to the press
   b. That Durant is an unfit nomination for Secretary of State
   c. It was Underwood who found Roy Kapeniak that eventually led to Kern's downfall
   d. That Underwood is using her and jeopardizing her position at work

12. What happens to Catherine as a result of Zoe Barnes releasing a story about Senator Catherine Durant's nomination as Secretary of State in place of Michael Kern?
    a. The release of Catherine's nomination drove the president to remove her from consideration
    b. Catherine becomes suspicious of Underwood's involvement in her nomination
    c. Catherine releases an official statement withdrawing her consideration for the position
    d. News surrounding her nomination led the President to consider her as candidate
13. Why was Michael Kern laughing on television significant?
    a. It suggested that Kern was not taking the interview seriously
    b. It suggested that Kern did not care about the Israeli-Palestinian conflict
    c. It showed that Kern had a sense of humor
    d. It showed the Kern would be a poor choice for Secretary of State
14. Why was it important for Frank Underwood to take care of matters in his home district instead of being on Capitol Hill working on the education bill?
    a. If Frank did not take care of matters in his home district, he may have not been re-elected
    b. Frank believed that consoling the family of the girl who passed away was a better use of his time than working on the education bill
    c. Frank need to go home every couple months in order for it to seem like he is serving his constituents
    d. If Frank did not go back to his home district, he would have been pressured into agreeing to add something to the bill that he disagreed with
15. Why did Frank Underwood say that the Walker administration's choice for Secretary of State was "excellent"?
    a. Frank thought the nominee was very qualified
    b. Frank was being polite, even though he was unhappy
    c. Frank was actually angry, and he uses the word excellent whenever he is angry
    d. Frank thought the nominee was under-qualified
16. With whom does Frank dance with at the Jefferson Ball?
    a. Catherine Durant
    b. Zoe Barnes
    c. Linda Vasquez
    d. Rachel Posner
17. Why was Zoe Barnes upset that her boss said she could not go on TV for a month?
    a. Zoe needed to go on TV to further her plan with Frank Underwood
    b. Zoe felt as if she was being treated like a child
    c. Doing interviews on television is Zoe's favorite activity to do

    d. Not going on television to do interviews prevented her from earning a large amount of money

18. Why did Peter Russo pretend to be talking on the phone to president-elect Garrett Walker?
    a. So the man Russo was meeting with would see him as important and powerful
    b. Russo wanted an excuse to talk dirty to Christina
    c. Russo thought Walker was on the phone, but it turned out to be Christina
    d. To stall in order to figure out what he was going to say to the man meeting with him
19. Why does Frank Underwood want Catherine Durant to be Secretary of State?
    a. Frank knows Durant will owe him favors in the future if he helps her get the position
    b. Frank thinks he can eventually replace Durant as Secretary of State
    c. Frank owes Durant a favor for a time she helped him pass a bill
    d. Frank think if Durant becomes Secretary if State she will be more likely to have an affair with him
20. What reasoning does Frank Underwood provide for killing the dog?
    a. It was painful and Frank felt bad
    b. Sometimes you have to do the unpleasant, necessary thing
    c. He has no patience for useless things like animals
    d. Pain is a sign of weakness
21. Why does Garrett Walker go back on his promise to Frank Underwood regarding the position Walker promised to give Underwood?
    a. Underwood is needed in Congress
    b. Walker decided to nominate Underwood for Speaker of the House instead
    c. Walker found out Underwood went behind his back to get the position
    d. Walker now believes Underwood is unfit for the position
22. What is the topic of the sermon the preacher delivers in the church in Washington D.C.?
    a. Faith
    b. Kindness
    c. Humility
    d. Trust
23. Why did Frank Underwood let Zoe Barnes into his home?
    a. Zoe showed Frank that she had a picture of him checking her out
    b. Zoe and Frank were having an affair
    c. Zoe and Frank were colleagues
    d. Zoe had recently received an email from Frank discussing recent legislation
24. What gift does Claire give Frank?
    a. A leg press machine
    b. A treadmill

c. A bench press
d. A rowing machine

25. What does Jeanine attribute Zoe's luck with groundbreaking stories to?
    a. That Zoe has a reliable inside source and she becomes envious
    b. She believes that Zoe is sleeping with somebody important to have that much luck
    c. She thinks that Zoe is blackmailing members of Congress for information
    d. That Zoe is using her sources to gather information and undermine her
26. What is the reason behind Claire firing half her staff?
    a. To make room for Gillian's organization and merge with World Well
    b. Claire found out that they were going behind her back to make changes to the charity
    c. Claire wants to make room for big corporation donors like SanCorp
    d. It was part of her plan with SanCorp to create new fracking locations overseas
27. What new finding does Frank Underwood explain to Oren Chase regarding Jessica Master's death?
    a. Jessica actually wasn't texting and driving
    b. Jessica would have survived the crash if there were guardrails
    c. That there was no way to prove that Jessica was texting and driving
    d. That Jessica was actually under the influence while driving
28. Why does Frank Underwood ask Jessica's father if he would like Frank to resign?
    a. To show humility, so Frank could get Jessica's father to do what Frank wants
    b. To show frustration, so Jessica's father would know Frank was inconvenienced by having to travel to South Carolina
    c. To show sarcasm, since Frank would never resign from Congress
    d. To show apathy, suggesting Frank really wanted to quit his job
29. What does Frank have Ed Meechum bring to Claire from Gaffney?
    a. White roses
    b. White tulips
    c. White daisies
    d. White carnations
30. Who does Peter Russo visit and then later do drugs with?
    a. A friend of Secretary of State nominee Michael Kern
    b. A man who worked with Secretary of State nominee Michael Kern on their college news paper
    c. A woman who was an old colleague of Secretary of State nominee Michael Kern
    d. A man who was a professor of Secretary of State nominee Michael Kern in college
31. Zoe Barnes is a reporter for which newspaper?

a. Washington Post
b. Washington Inquirer
c. Washington Herald
d. Washington Times

32. Who does Frank Underwood want to replace Michael Kern with a nominee for Secretary of State?
    a. Christina Gallagher
    b. Catherine Durant
    c. Linda Vasquez
    d. Rachel Posner
33. What elected position does Frank Underwood hold?
    a. Congressman
    b. Senator
    c. Governor
    d. Vice President
34. With how many days does the Walker administration want the education bill on the floor of Congress?
    a. Within the first fifty days of being in office
    b. Within the first one-hundred days of being in office
    c. Within the first two-hundred days of being in office
    d. Within the first thirty days of being in office
35. Frank Underwood is originally from which state?
    a. North Carolina
    b. Georgia
    c. South Carolina
    d. Alabama
36. What is Linda Vasquez's role in the White House?
37. What is the name of Frank Underwood's Chief of Staff and right-hand man?
38. With what criminal offense was Peter Russo charged?
39. What D.C. BBQ location does Frank Underwood frequent?
40. Why was Evelyn Baxter upset that Claire's charity was making budget cuts?

Appendix B

Table B1

BWEA Item Analysis (Q1-30)

Item #	Standard Deviation	Item Difficulty	Item Discrimination	Alpha (.7) if Item Deleted	Comments
Q1	0.480	0.647	0.085	0.705	Alpha improves if item is deleted. Item has weak discrimination. Item is performing poorly.
Q2	0.499	0.547	0.271	0.689	Item is performing adequately.
Q3	0.468	0.680	0.211	0.694	Item is performing adequately.
Q4	0.391	0.813	0.232	0.692	Item may be too easy. Item is performing adequately.
Q5	0.225	0.947	0.139	0.698	Item may be too easy. Item is performing adequately.
Q6	0.484	0.633	0.277	0.688	Item is performing adequately.
Q7	0.368	0.840	0.249	0.691	Item may be too easy. Item is performing adequately.
Q8	0.355	0.853	0.259	0.691	Item may be too easy. Item is performing adequately.
Q9	0.454	0.713	0.475	0.672	Item has strong discrimination. Item is performing well.
Q10	0.310	0.893	0.405	0.683	Item may be too easy. Item has strong discrimination. Item is performing well.
Q11	0.473	0.667	0.403	0.678	Item is performing adequately.

Q12	0.301	0.900	0.412	0.683	Item may be too easy. Item has strong discrimination. Item is performing well.
Q13	0.482	0.640	0.097	0.704	Alpha improves if item is deleted. Item has weak discrimination. Item is performing poorly.
Q14	0.368	0.840	0.309	0.687	Item may be too easy. Item is performing adequately.
Q15	0.497	0.567	0.126	0.702	Alpha improves if item is deleted. Item is performing poorly.
Q17	0.406	0.793	0.393	0.680	Item is performing adequately.
Q18	0.502	0.500	0.092	0.705	Alpha improves if item is deleted. Item has weak discrimination. Item is performing poorly.
Q19	0.468	0.680	0.243	0.691	Item is performing adequately.
Q20	0.348	0.860	0.202	0.694	Item may be too easy. Item is performing adequately.
Q21	0.429	0.760	0.378	0.681	Item is performing adequately.
Q22	0.480	0.353	0.243	0.691	Item is performing adequately.
Q23	0.212	0.953	0.068	0.700	Item may be too easy. Item has weak discrimination. Item is performing poorly.
Q24	0.197	0.960	0.023	0.702	Alpha improves if item is deleted. Item may be too easy. Item has weak discrimination. Item is performing poorly.
Q25	0.272	0.920	0.173	0.696	Item may be too easy. Item is performing adequately.
Q26	0.471	0.673	0.326	0.684	Item is performing adequately.

Q27	0.180	0.967	0.234	0.695	Item may be too easy. Item is performing adequately.
Q28	0.326	0.880	0.247	0.692	Item may be too easy. Item is performing adequately.
Q29	0.368	0.840	0.299	0.688	Item may be too easy. Item is performing adequately.
Q30	0.348	0.860	0.140	0.698	Item may be too easy. Item is performing adequately.

Appendix C

Participant Instructions (Partial)

Spaced Condition 1

"You are about to watch episode 1 of House of Cards, a Netflix original series. This episode is roughly 45 minutes long. You will eventually be quizzed on the information presented in the show, so please do your best to pay attention. You are not permitted to take notes while viewing any of the episodes. As a reminder – you may withdraw your participation from this experiment at any time and leave the room.

Massed Condition 1

"You are about to watch episodes 1, 2, and 3 of House of Cards, a Netflix original series. Each episode is roughly 45 minutes long, so you will be here for 2-3 hours. If you need to use the bathroom, or fill up your water bottle, please do so now. We'll give everyone 5 minutes to do so before we continue." -- PAUSE FOR 5 MINUTES --"You will eventually be quizzed on the information presented in the show, so please do your best to pay attention. You are not permitted to take notes while viewing any of the episodes. As a reminder – you may withdraw your participation from this experiment at any time and leave the room. We will also have a short, 5 minute break after the completion of episode 2."

Spaced/Massed Quiz

You have watched 3 episodes of House of Cards in (a) previous session(s). Today, you will take a quiz to gauge your memory of the show's content. Once you finish the quiz, you are free to leave and your participation in this study will be complete. This quiz should take about 20 minutes. Please take your time and mark the best answer. Please do not look up any answers or consult with each other for help. Raise your hand if you have any questions. You may click on the arrow to begin."

Appendix D

Normality Tables

Table D1

Normality of BWEA Total Recognition Memory Score (0-35) for each Viewing and Delay Group

viewing_condition	Delay	Shapiro Wilk *W*	*Df*	*P*
Spaced	One-Week	.926	29	.043
	Four-Week	.949	20	.359
Massed	One-Week	.883	20	.020
	Four-Week	.960	23	.463

Table D2

Normality of BWEA Recurring Question Score (0-7) for each Viewing and Delay Group

viewing_condition	delay	Shapiro Wilk *W*	*df*	*P*
Spaced	One-Week	.872	29	.002
	Four-Week	.912	20	.070
Massed	One-Week	.931	20	.161
	Four-Week	.936	23	.144

Table D3

Normality of BWEA Total Recall Memory Score (0-5) for each Viewing and Delay Group

viewing_condition	delay	Shapiro Wilk *W*	*df*	*P*
Spaced	One-Week	.835	29	<.001
	Four-Week	.908	20	.059
Massed	One-Week	.943	20	.276
	Four-Week	.914	23	.049

Table D4

Normality of BWEA Episode 3 Memory Score (0-5) for each Viewing and Delay Group

viewing_condition	delay	Shapiro Wilk *W*	*df*	*p*
Spaced	One-Week	.657	29	<.001
	Four-Week	.923	20	.112
Massed	One-Week	.830	20	.002
	Four-Week	.940	23	.002

References

Ambridge B., Theakston A. L., Lieven E. V., Tomasello M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Journal of Cognitive Development*, *21*, 174–193. Doi: 10.1016/j.cogdev.2005.09.003.

Anghelcev, G., Sar S., Martin J. D., & Moultrie, J. L. (2021). Binge-watching serial video content: Exploring the subjective phenomenology of the binge-watching experience. *Mass Communication and Society*, *24*(1), 130-154. https://doi.org/10.1080/15205436.2020.1811346

Austin, M., & Shapovalov, Y. (2020). Psychometric data analysis of the Binge-Watching Episode Assessment [unpublished manuscript]. Department of Psychology, James Madison University, Harrisonburg, U.S.

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A Metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52*(4), 566-577. doi:10.1016/j.jml.2005.01.012

Bakrania, K., Edwardson, C., Khunti, K., Bandelow, S., Davies, M. J., & Yates, T. (2018). Associations between sedentary behaviors and cognitive function: Cross-sectional and prospective findings from the uk biobank. *American Journal of Epidemiology, 187*(3), 441–454. https://doi.org/10.1093/aje/kwx273

Bell, M. C., Kawadri, N., Simone, P. M., & Wiseheart, M. (2014). Long-term memory, sleep, and the spacing effect. *Memory*, *22*(3), 276-283. https://doi.org/10.1080/09658211.2013.778294

Benjamin A. S., Tullis J. (2010). What makes distributed practice effective? *Journal of*

Cognitive Psychology, 61, 228–247. Doi: 10.1016/j.cogpsych.2010.05.004

Bjork R. A., Allen T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, 9, 567–572. Doi: 10.1016/S0022-5371(70)80103-7.

Bjork R. A., Bjork E. L. (1992). "A new theory of disuse and an old theory of stimulus fluctuation," in From Learning Processes to Cognitive Processes: Essays in Honor of William eds Estes K., Healy A., Kosslyn A., Shiffrin R. (Hillsdale, NJ: Erlbaum;) 35-67.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory,* 85-123. Washington: Winston, 1972.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness of hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547-1522. DOI:10.1037/a0024140.

Castro, D., Rigby, J.M., Cabral, D., Nisi, V. (2019). The binge-watcher's journey: Investigating motivations, contexts, and affective states surrounding Netflix viewing. *Convergence: The International Journal of Research into New Media Technology*, 27(1), 3-20.

Cepeda, N. J., Pashler H., Vul E., Wixted J. T, & Rohrer D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. Psychological Bulletin, 132, 354–380.

Cepeda N. J., Vul E., Rohrer D., Wixted J. T., Pashler H. (2008). Spacing effects in

learning a temporal ridgeline of optimal retention. *Psychological Science, 19,* 1095–1102. Doi: 10.1111/j.1467-9280.2008.02209.x.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56*(4), 236-246. DOI: 10.1027/1618-3169.56.4.236.

Conlin, L.T. (2015). *There goes the weekend: Understanding television binge-watching.* [Unpublished doctoral dissertation]. The University of Alabama.

Conlin, L., Billings, A.C., Averset, L. (2016). Time-shifting vs. appointment viewing: The role of fear of missing out within TV consumption behaviours. *Communication & Society, 29*(4), 151–164. https://doi.org/10.15581/003.29.4.151-164

Dail T. K., Christina R. W. (2004). Distribution of practice and metacognition in learning and long-term retention of a discrete motor task. *Research Quarterly for Exercise and Sport, 75*, 148–155. Doi: 10.1080/02701367.2004.10609146.

Dal Cin, S., Zanna, M. P., & Fong, G. T. (2004). Narrative persuasion and overcoming resistance. In E. S. Knowles & J. A. Linn (Eds.), *Resistance and persuasion*. Mahweh, NJ: Lawrence Erlbaum Associates Publishers.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS*

Catalog of Selected Documents in Psychology, 10, 85.

Delaney P. F., Verkoeijen P. P. & Spirgel A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. *Psychology of*

Learning and Motivation, *53*, 63–147. Doi: 10.1016/S0079-7421(10)53003-2

Deloitte. (2017). Digital democracy survey: A multi-generational view of consumer technology, media, and telecom trends. In *Digital Democracy Survey*, 11th ed. Oakland: Deloitte Development LLC.

Dempster F. N., (1988). The spacing effect: A case study in the failure to apply the results of psychological research. American Psychologist, 43, 627–634.

Dunlosky J., Rawson K. A., Marsh E. J., Nathan M. J. & Willingham D. T. (2014). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. DOI: 10.1177/1529100612453266.

Ebbinghaus, H. (1885). Über das Gedchtnis. Untersuchungen zur experimentellen Psychologie. Memory. A Contribution to Experimental Psychology. New York: Teachers College, Columbia University (Reprinted Bristol: Thoemmes Press, 1999).

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program.

Behavior Research Methods, Instruments, & Computers, 28(1), 1-11. https://doi.org/10.3758/BF03203630

Exelmans, L. & Van den Bulck, J. (2017). Binge-viewing, sleep, and the role of pre-sleep arousal. *Journal of Clinical Sleep Medicine, 13*(8), 1001–1008. https://doi.org/10.5664/jcsm.6704

Fancourt, D., & Steptoe, A. (2019). Television viewing and cognitive decline in older age: Findings from the english longitudinal study of ageing. *Scientific Reports, 9*(1), 1-8. https://doi.org/10.1038/s41598-019-39354-4

Flayelle, M., Maurage, P., Karila, L., Claus, V., & Billieux, J. (2019). Overcoming the unitary exploration of binge-watching: A cluster analytical approach. Journal of Behavioral Addiction, 8(3), 586–602. https://doi.org/10.1556/2006.8.2019.53

Fogler, K., Giordano, J. P., Chusid, S. A., & Roberts, C. (2017). The spacing effect and binge-watching [unpublished manuscript]. Department of Psychology, James Madison University, Harrisonburg, U.S.

Gerrig, R. J. (1993). *Experiencing narrative worlds: On the psychological activities of reading*.

New Haven, CT: Yale University Press.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95–112.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public

narratives. *Journal of Personality and Social Psychology, 79*(5), 701-721. https://doi.org/10.1037/0022-3514.79.5.701

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371–377.

Greeno, J. G. (1964). Paired-associated learning with massed and spaced repetitions of Items. *Journal of Experimental Psychology, 67*, 286-295.

Hamer, M. & Stamatakis, E. (2014). Prospective study of sedentary behavior, risk of depression, and cognitive impairment. Medicine and Science in Sports and Exercise, 46(4), 718–723. doi: 10.1249/MSS.0000000000000156

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), Theories in cognitive psychology: The Loyola symposium (pp. 77–97). Potomac, MD: Erlbaum.

Harris Interactive (April 8, 2013). *Americans Talking Advantage of Ability to Watch TV on Their Own Schedules*. The Harris Poll. https://theharrispoll.com/new-york-n-y-april-8-2013-video-on-demand-dvrs-streaming-content-entire-seasons-of-television-shows-collected-in-dvd-box-sets-americans-tv-viewing-habits-were-once-at-the-mercy-of-network/

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior*, *12*, 229-238.

Hintzman, D. L., & Block, R. A. (1970). Memory judgements and the effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *9*, 561-566.

Hoang, T. D., Reis, J., Zhu, N., Jacobs Jr. D. R., Launer, L. J., Whitmer, R. A., Sidney, S., & Yaffe, K. (2015). Effect of early adult patterns of physical activity and television viewing on midlife cognitive function. *JAMA Psychiatry*, *73*(1) 73-79. doi:10.1001/jamapsychiatry.2015.2468

Hopkins R. F., Lyle K. B., Hieb J. L., & Ralston P. A. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educ Psychol Rev*, *28*, 853-873. DOI 10.1007/s10648-015-9349-8.

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30,* 138–149. DOI: 0093-5301/2004/3001-0011$10.00.

Jenner, M. (2014). Is this tviv? On netflix, tviii and binge-watching. New Media & Society, 18(2), 257-273. https://doi.org/10.1177/1461444814541523

Kesse-Guyot, E., Charreire, H., Andreeva, V. A., Touvier, M., Hercberg, S., Galan, P., & Oppert, J. M. (2012). Cross-sectional and longitudinal associations of different sedentary behaviors with cognitive performance in older adults. *Plos One, 7*(10), 1-10. https://doi.org/10.1371/journal.pone.0047831

Lehmann H., Sparks F. T., Spanswick S. C., Hadikin C., McDonald R. J., Sutherland R. J. (2009). Making context memories independent of the hippocampus. *Learning and Memory, 16,* 417–420. Doi: 10.1101/lm.1385409.

Maddox G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28,* 684–706. Doi: 10.1080/20445911.2016.1181637.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 828-835.

McNamara, M. (2012, Jan. 15). Critic's notebook: The side effects of binge television. *Los Angeles Times*. https://www.latimes.com/archives/la-xpm-2012-jan-15-la-ca-netflix-essay-20120115-story.html

Melton, A. W. (1967). Repetition and retrieval from memory. *Science, 158,* 532.

Melton, A.W. (1970). The situation with respect to the spacing of repetitions and

memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 596-606.

Panda, S. & Pandey, S. C. (2017). Binge-watching and college students: Motivations and outcomes. *Young Consumers, 18*(4), 425–438. https://doi.org/10.1108/YC-07-2017-00707

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., and Metcalfe, J. (2007) *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. retrieved from http://ncer.ed.gov.

Peterson, T. G. (2016). To binge or not to binge: A Qualitative analysis of college students' binge watching habits. *The Florida Communication Journal*, *44*(1), 77-87.

Pittman, M. & Steiner, E. (2019). Transportation or narrative completion? Attentiveness during binge-watching moderates regret. *Social Science*, *8*(3), 99-113. https://doi.org/10.3390/socsci8030099

Raaijmakers J. G. (2003). Spacing and repetition effects in human memory: application of the SAM model. *Cognitive Science, 27*, 431–452. doi: 10.1207/s15516709cog2703_5.

Riddle, K., Peebles, A., Davis, C., Xu, F., & Schroeder, E. (2018). The addictive potential of television binge watching: Comparing intentional and unintentional binges. *Psychology of Popular Media Culture*, *7*(4), 589-604. https://doi.org/10.1037/ppm0000167

Rohrer D., Taylor K. (2007). The shuffling of mathematics problems improves

learning. *Instructional Science, 35*, 481–498.
Doi: 10.1007/s11251-007-9015-8.

Rubenking, B. & Bracken, C. C. (2018). Binge-watching: A suspenseful, emotional habit. *Communication Research Reports, 35*(5), 1–11.
https://doi.org/10.1080/08824096.2018.1525346

Shaughnessy J. J., Zimmerman J., & Underwood B. J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *11*, 1-12.

Shebilske W. L., Goettl B. P., Corrington K., Day E. A. (1999). Interlesson spacing and task-related processing during complex skill acquisition. *Journal of Experimental Psychology: Applied, 5*, 413. doi: 10.1037/1076-898x.5.4.413.

Simmons A. L. (2011). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education, 59,* 357–368. doi: 10.1177/0022429411424798.

Smith C. D., Scarf D. (2017). Spacing repetitions over long timescales: a review and a reconsolidation explanation. *Frontiers in Psychology*, *8*, 962. doi:10.3389/fpsyg.2017.00962.

Spruance, L., Karmakar, M., Kruger, J., & Vaterlaus, J. (2017). "Are you still watching?" Correlations between binge TV watching, diet, and physical activity. Journal of Obesity and Weight Management, 1–8.

Sobel H. S., Cepeda N. J., & Kapler I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology*, *25*, 763-767. doi: 10.1002/acp.1747

Starosta, J., Izydorczyk, B., Lizińczyk, S. (2019). Characteristics of people's binge-watching behavior in the "entering into early adulthood" period of life. Health Psychology Report, 7(2), 149–164. https://doi.org/10.5114/hpr.2019.83025

Starosta, J. A., & Izydorczyk, B. (2020). Understanding the phenomenon of binge-watching – a systematic review. *International Journal of Environmental Research and Public Health*, *17*(12), *4469*. https://doi.org/10.3390/ijerph17124469

Steiner, E. (2018). *Binge-watching killed the idiot box: The changing identities of viewers and television in the experiential, streaming video age* (Publication No. 10813061) [Doctoral Dissertation, Temple University]. ProQuest Dissertations Publishing.

Steiner, E., & Xu, K. (2018). Binge-watching motivates change: Uses and gratifications of streaming video viewers challenge traditional TV research. *Convergence: The International Journal of Research into New Media Technologies*, *26*(1), 82-101. https://doi.org/10.1177/1354856517750365

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*, 529–536.

Vander Linde E., Morrongiello B. A., Rovee-Collier C. (1985). Determinants of retention in 8-week-old infants. *Developmental Psychology, 21,* 601–613. Doi: 10.1037/0012-1649.21.4.601.

Wang, J. Y. J., Zhou, D. H. D., Zhang, M., Deng, J., Tang, M., Gao, C., Li, J., Lian, Y., & Chen, M. (2006). Leisure activity and risk of cognitive impairment: The Chongqing aging study. *Neurology, 66*(6), 911–913. https://doi.org/10.1212/01.wnl.0000192165.99963.2a